DOCUMENT RESUME

ED 426 085 TM 029 313

AUTHOR Roberts, J. Kyle
TITLE Thurstone's Method of Equal-Appearing Intervals in Measuring Attitudes: An Old Method That Is Not Forgotten.
PUB DATE 1998-11-00
NOTE 15p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (New Orleans, LA, November 4-6, 1998).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Attitude Measures; Elementary Secondary Education; Ethics; Personality Assessment; Program Evaluation; *Scaling; *Test Construction; *Values
IDENTIFIERS *Character Education; Equal Interval Scoring; *Thurstone Model

ABSTRACT

Many school districts face the problem of evaluating new programs to train students in ethics and moral decision making. Using conventional personality tests in program evaluation may be helpful, but probably will not provide measures for the attitudes that are targeted by the intervention. The method of equal- appearing intervals developed by L. Thurstone (1959) provides a viable solution for this problem. How school districts could use this type of scaling to measure attitudes specific to the design of character education programs is discussed. The relative simplicity of Thurstone scaling means that most schools could develop and administer this method in-house. The relative ease of construction would allow school districts not only to measure the effectiveness of an intervention but also to identify children who might be best suited for character education programs. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 426 085

Running head: THURSTONE'S METHOD IN MEASURING ATTITUDES

Thurstone's Method of Equal-Appearing Intervals in Measuring Attitudes: An Old Method that is not Forgotten

J. Kyle Roberts
Texas A&M University 77843-4225

J. Kyle Roberts

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM029313

Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans, November 4, 1998.

Abstract

Many school districts are facing problems of what to do with new programs to train ethics and moral decision making to students. Because of the growing concern from parents and administrators, many such programs have made their way into the public school setting. One of the problems with these interventions is that the ineffectiveness often goes unmeasured. Employing the use of conventional personality tests in program evaluation may be helpful, but probably will not provide measures for the attitudes that are targeted by the intervention. Thurstone's method of equal-appearing intervals provides a viable alternative to this problem. This paper shows how school districts could use this type of scaling to measure attitudes specific to the design of character education programs.

In recent years, there has been a rise in what professionals call a need for "character education" in the public schools. This need has arisen partly out of the seemingly endless onslaught of morally reprehensible acts in public schools. What was once termed a need to educate children on the dangers of drugs has now become a need to educate them on the dangers of their own morality. Although not proven as of yet, some professionals believe that there is a need for a part of the educational curriculum to be devoted to the development of the child's morality (Goodman, 1997).

In a seeming haphazard way, some of these school districts have sought curriculum help from counselors who actually have no training in character education. Some school districts have even been taken to court for implementing programs that are religious in nature, such as the program started by Bill Gothard (Gothard, 1998). Although some of these programs may be successful in developing a sense of morality in students who may be targeted as potentially at-risk, there has been relatively little work done to provide a quantitative measure of the effectiveness or ineffectiveness of such programs.

There exists a measure called "process tracing" which was developed in an attempt to provide information measuring attitudes for which there are no pre-existing instruments. This type of measure is largely modeled after the process developed by Lawrence Kohlberg when analyzing his "Heinz Dilemma" and other moral judgement case studies (Woolfolk, 1998). Although this type of measure provides a thorough investigation into the thought processes of the adolescent, this is a lengthy process that is largely qualitative in nature. The process also requires measurement by someone who has had extensive training; this could be potentially costly to a school district.

It may be proposed, also, that single-case research design (Kazdin, 1982) could be a possible way to measure students' attitudes in such programs. If the character education intervention programs were designed to change behaviors, then a single-case research design would provide a powerful way to measure individual student outcomes on observed or measured variables. The difficulty, however, is that most character education programs are designed to produce a change in attitudes rather than a change in behavior. In many cases, though, it is believed that such character education programs produce a change in attitude that will ultimately produce a change in behavior. If this is the case, then a single-case research design would be useful in helping to strengthen the measured effectiveness of character education programs.

Let us suppose, for example, that a student is referred to a character education program because he or she has been caught stealing from other students. The goal of this program may be not only to keep this student from stealing again, but to teach the student both the moral reprehensibility of stealing and respect for other people's property. Single –case research could be used to measure whether or not a behavior change occurred, but it would be necessary to use other means to measure whether or not there was a change in attitude.

One of the difficulties of programs that teach ethics and morality is that they are often evaluated just by means of behavioral outcomes. When measuring scholastic ability, one need merely look at outcome scores on standardized tests. But when measuring moral and ethical judgement, no hard and fast quantitative rubric has been developed. And even if a standardized rubric was developed, there would probably be a lot of discrepancy concerning the anchors that the creators of this instrument chose.

What is now needed is an instrument that is relatively simple to use that provides reliable and valid results. Because of the variability of these character education programs implemented in schools, it is also necessary that such a measure of attitudes have the ability to be developed "in house" so that administrators can identify potentially at-risk students and measure outcomes on specific criteria. Thurstone's method of equal-appearing intervals could provide school districts with a reliable and valid answer to these problems. By allowing schools to create their own instruments, these scaling methods could help to strengthen character training programs by allowing the measurement of the attitudes specifically targeted by the intervention.

Thurstone's Scale

Thurstone postulated that for any psychological object: "(1) our reactions to such stimuli were subjective; and (2) our judgement or preference for an object may vary from instance to instance" (Dunn-Rankin, 1983). Thurstone also noted that although we often have particular reactions to differing stimuli in differing circumstances, there is still a most common reaction which he called the "modal reaction" (Dunn-Rankin, 1983). Because of this reaction, he was able to develop his method of equal-appearing intervals to measure preference in attitudes.

For the purpose of illustration a heuristic example is employed to demonstrate the development of this type of Thurstone scale. Let us suppose that we are developing an intervention to work with students who may be at-risk for stealing. Our first task in this intervention would be to identify those students who would benefit the most from such a program. This can be done either by referral or by administering an instrument to measure attitudes.

The first step in developing such an instrument is to gather opinions about the subject to be studied--in this case, stealing. These opinions should be couched in the form of statements ranging from decidedly in favor of stealing through neutral statements about stealing to those that are very much opposed to stealing (Thurstone, 1959). In the original sample this particular scale included over 100 statements.

These statements should then be distributed to a number of individuals who have an understanding of the scaling method. Each of these 100 statements were typed out and distributed to the individuals conducting the research in an effort to eliminate the most unsatisfactory statements (Thurstone, 1959). After this was done, 37 statements remained that seemed representative of the larger sample, as illustrated in Table 1.

INSERT TABLE 1 ABOUT HERE

Once this smaller sample is developed, the next step is to scale the items that are to be part of the instrument. To do this, it is necessary to gather "judges" who will rate each item. The rating scale chosen for this instrument was a 1-to-7 scale with the anchors being "extremely favorable attitude towards stealing" and "extremely unfavorable attitude towards stealing", respectively. When having the judges rate each of the 37 statements, it is important to have them rate the favorableness or unfavorableness of each item. You do NOT want the judges to give their attitude towards stealing or whether or not they agree or disagree with the statement (Trochim, 1998). It is also important that the judges not assume any particular distribution of the statements (Torgerson, 1958). The judges were prompted for each statement with the question, "How strongly in favor of stealing is someone who endorses this statement?".

The next step is to compute both the median and interquartile range for the ratings for each of the 37 statements. In Table 2, we can see a histogram of scores gathered from the judges on statement number 10. The median value of scores for each of the statements is that value above and below which 50% of the ratings fall. The interquartile range is the difference between the third and first quartile, or Q3-Q1, and is a measure of the ambiguity of each statement (Thurstone, 1959). The first and third quartiles are found by computing the value below which 25% of the scores fall and below which 75% of the scores fall, respectively. Table 3 illustrates the findings from each of the 37 statements.

INSERT FIGURE 1 AND TABLE 2 ABOUT HERE

After this is done, final statements must be selected for the instrument. In doing this, it is important to select statements that strongly represent each of the scaling points. To do this, items were selected within each value that had the lowest interquartile range, thus yielding the least amount of ambiguity and variability across judges. For example, as reflected in Table 3, four items (i.e., 1, 7, 9, and 16) are eligible to represent the scale value of "4", because they have medians of 4. However, only items 1 and 9 have interquartile ranges (i.e., 75%ile – 25%ile) of 1. Item 16 would have been least satisfactory, because its interquartile range was largest (i.e., 3).

In the case of multiple items with the same median and interquartile range, judgement should be used to select statements that are most easily understood. This complies with the three criteria that Thurstone mentioned for selecting statements for the final instrument (Thurstone, 1959). From Table 3, items 1, 2, 4, 5, 9, 10, 11, 13, 14, 17, 19, 20, 25, 26, 28, and 30 were chosen to be representative of each scale value.

Once these items have been placed in random order, the instrument is ready to be administered. When administered, the student should be asked to either agree or disagree with each of the statements that they are given. Any items left blank should be treated as statements with which the student disagrees (Trochim, 1988). Once all statements are marked, the student's score is computed by simply averaging the median values of all of the statements with which the student agreed. Hence, a student with a small score has a more favorable attitude towards stealing than someone who has a large score.

Table 4 illustrates one person's possible score on this instrument. In this case, the student agreed with statements 1, 9, 10, 11, 13, 14, and 25. By averaging the median values for each of these statements, the student yields an overall score on the stealing instrument of 5.29. This score means that this student's attitude is less positive toward stealing than any student who scores lower on the scale. This is somewhat important, because we cannot, as Thurstone (1959) notes, say that one score is better or worse than another.

INSERT TABLE 3 ABOUT HERE

Summary

The relative simplicity of Thurstone scaling is the reason that it was chosen as a means by which to measure attitudes of students. This a method that most schools could develop and administer in-house. The relative ease of construction would allow school districts to measure not only the effectiveness of an intervention, but also help identify those children who might be best suited for character education programs. This method

of scaling also produces an instrument that is easily interpreted and thus can be administered by most anyone.

References

Dunn-Rankin, P. (1983). Scaling methods. Mahwah, N J. Erlbaum.

Goodman, J. F. (1997, September 24). Talk of the good is good talk. Education week (pp. 32-35).

Gothard, W. (1998). Character first education. In Institute in basic life principles [On-line]. Available: http://www.iblp.org/service.asp

Kazdin, A. E. (1982). Single-case research design: Methods for clinical and applied settings. Oxford: Oxford University Press.

Thurstone, L. L. (1959). The measurement of values. Chicago: University of Chicago Press.

Torgerson, W. (1958). Theory and methods of scaling. New York: Wiley.

Trochim, B. (1998). Thurstone scaling. In The knowledge base: An online research methods textbook [On-line]. Available: http://trochim.human.cornell.edu/kb/scalthur.htm

Woolfolk, A. E. (1998). Educational psychology. Needham Heights, MA. Allyn and Bacon.

Table 1

Statements Concerning Stealing

1. Stealing for the purpose of helping someone else is okay
2. It is okay to steal from wealthy people
3. If it will get you out of trouble, it is okay to steal
4. Stealing is always wrong
5. Stealing is rarely ever wrong
6. It is okay to steal from people younger than you
7. It is okay to steal if it will not hurt anyone
8. It is okay to steal if the item costs less than $5.00
9. It is okay to steal if the item costs less than $1.00
10. Stealing for the purpose of staying alive is okay
11. Stealing in order to protect yourself is okay
12. I hate people who steal from me
13. People who steal are bad people
14. I would never steal from a friend
15. I would never steal for myself, only to help others
16. Sentencing should be lighter for people who steal than for people who drive drunk
17. Stealing really hurts no one
18. I would steal food to feed myself
19. I would steal if I knew that I would not get caught
20. I would steal something for a friend
21. I would steal if I didn't have enough money to buy someone a birthday/Christmas present
22. I would steal from a criminal
23. I would steal an animal if I didn't have anything to eat
24. People who steal cars should be punished severely
25. People who steal food should not be punished severely
26. People steal because they are poor
27. People steal because they are criminals
28. People who steal should go to prison
29. Stealing is only bad if you do it all of the time
30. Stealing once in a while is okay to do
31. Even if someone only stole something once, they are still a thief
32. Thieves are people who steal all of the time
33. People who only steal some of the time are not thieves
34. I would never steal anything
35. Stealing a pencil from the school is okay if you are going to use it
36. I would steal from my parents
37. It is okay to steal from people you don't know

Figure 1
Histogram of Scores on Statement 10

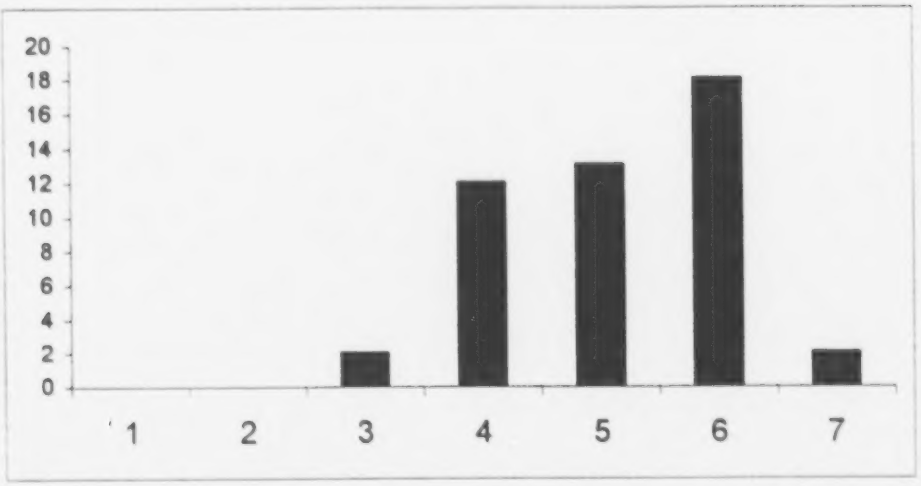

Table 2
List of Opinions on the 37 Statements About Stealing

Statement	Mdn	Q3	Q1	I-Range
1 Stealing for the purpose of helping someone else is okay	4	4	3	1
2 It is okay to steal from wealthy people	2	2	2	0
3 If it will get you out of trouble, it is okay to steal	2	4	2	2
4 Stealing is always wrong	7	7	7	0
5 Stealing is rarely ever wrong	2	2	1	1
6 It is okay to steal from people younger than you	2	4	2	2
7 It is okay to steal if it will not hurt anyone	4	4	2	2
8 It is okay to steal if the item costs less than $5.00	3	3	2	1
9 It is okay to steal if the item costs less than $1.00	4	4	3	1
10 Stealing for the purpose of staying alive is okay	6	6	4	2
11 Stealing in order to protect yourself is okay	6	6	4	2
12 I hate people who steal from me	7	7	5	2
13 People who steal are bad people	7	7	7	0
14 I would never steal from a friend	5	6	4	2
15 I would never steal for myself, only to help others	5	6	3	3
16 Sentencing should be lighter for people who steal than for people who drive drunk	4	6	3	3
17 Stealing really hurts no one	1	2	1	1
18 I would steal food to feed myself	6	6	3	3
19 I would steal if I knew that I would not get caught	2	3	2	1
20 I would steal something for a friend	3	4	2	2
21 I would steal if I didn't have enough money to buy someone a birthday/Christmas present	2	5	1	4
22 I would steal from a criminal	3	5	2	3
23 I would steal an animal if I didn't have anything to eat	5	6	5	1
24 People who steal cars should be punished severely	5	6	3	3
25 People who steal food should not be punished severely	5	6	4	2
26 People steal because they are poor	3	4	2	2
27 People steal because they are criminals	7	7	7	0
28 People who steal should go to prison	7	7	6	1
29 Stealing is only bad if you do it all of the time	2	4	2	2
30 Stealing once in a while is okay to do	3	4	2	2
31 Even if someone only stole once, they are still a thief	7	7	7	0
32 Thieves are people who steal all of the time	5	6	2	4
33 People who only steal some of the time are not thieves	2	3	2	1
34 I would never steal anything	7	7	7	0
35 Stealing a pencil from the school is okay if you are going to use it	3	5	2	3
36 I would steal from my parents	2	3	1	2
37 It is okay to steal from people you don't know	2	2	2	0

Note: The interquartile range (I-Range) is Q3 – Q1 (e.g., 4 – 3 = 1).

Table 3
A Possible Score on the Instrument

Statement	Agree	Disagree	Mdn
Stealing for the purpose of helping someone else is okay	X		4
Stealing once in a while is okay to do		X	-
It is okay to steal from wealthy people		X	-
Stealing is rarely ever wrong		X	-
It is okay to steal if the item costs less than $1.00	X		4
Stealing in order to protect yourself is okay	X		6
People who steal are bad people	X		7
I would never steal from a friend	X		5
People who steal should go to prison		X	-
Stealing really hurts no one		X	-
I would steal if I knew that I would not get caught		X	-
I would steal something for a friend		X	-
Stealing is always wrong		X	-
People who steal food should not be punished severely	X		5
People steal because they are poor		X	-
Stealing for the purpose of staying alive is okay	X		6

Mean = 5.29